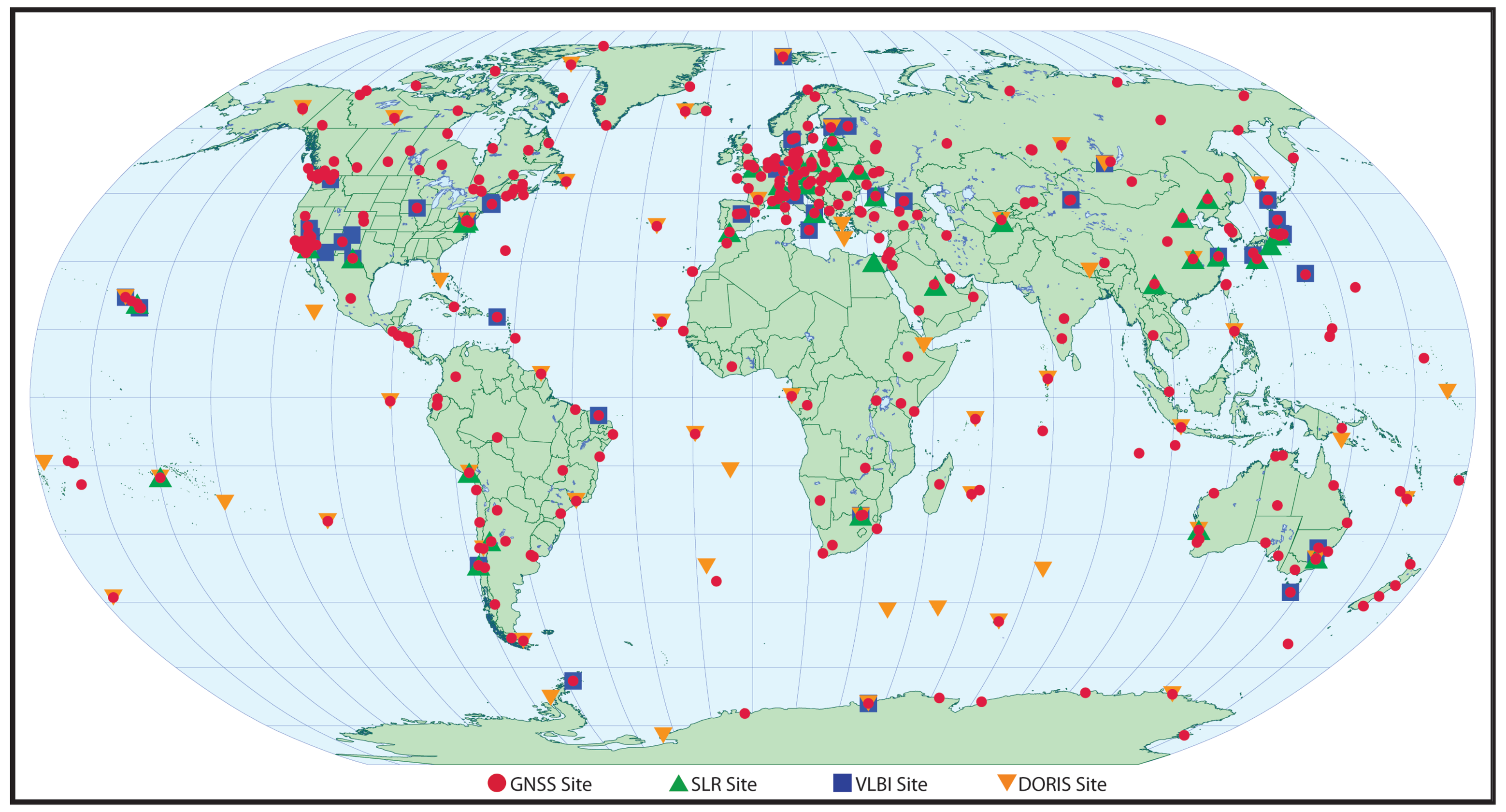
RECENT DEVELOPMENTS AT THE CDDIS

Carey Noll, NASA GSFC
Maurice Dube, NASA GSFC/SSAI
Patrick Michael, NASA GSFC/Catholic University

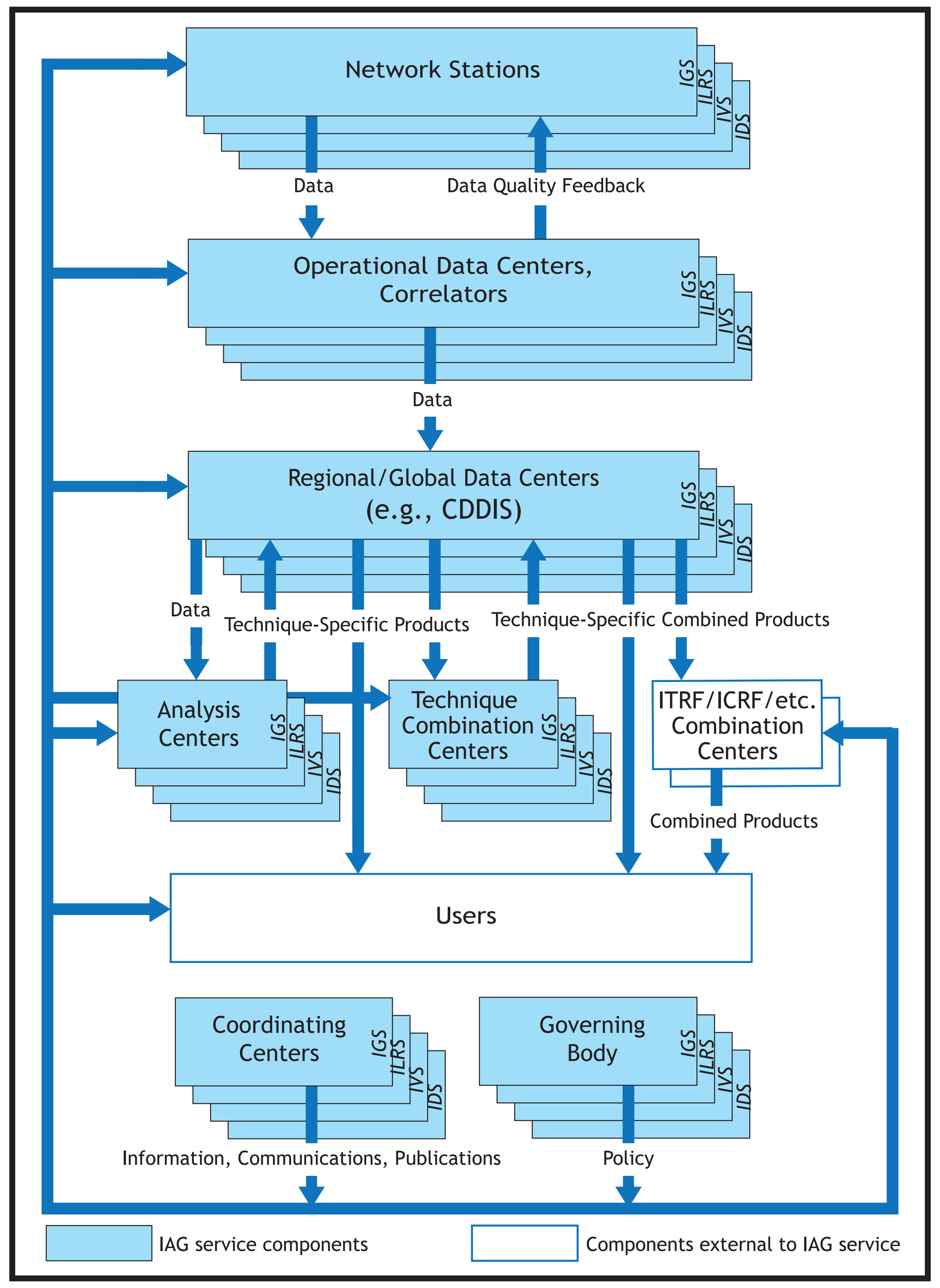
ABSTRACT

Since 1982, the Crustal Dynamics Data Information System (CDDIS) has supported the archive and distribution of geodetic data products acquired by NASA as well as national and international programs. These data include GNSS (Global Navigation Satellite System), SLR and LLR (Satellite and Lunar Laser Ranging), VLBI (Very Long Baseline Interferometry) and DORIS (Doppler Orbitography and Radiolocation Integrated by Satellite). The CDDIS data system and its archive have become increasingly important to many national and international science communities, particularly several of the operational services within the International Association of Geodesy (IAG) and its project the Global Geodetic Observing System (GGOS), including the International DORIS Service (IDS), the International GNSS Service (IGS), the International Laser Ranging Service (ILRS), the International VLBI Service for Geodesy and Astrometry (IVS), and the International Earth Rotation Service (IERS). The CDDIS has recently implemented a new, distributed hardware architecture. This poster will include background information about the system and its user communities, archive contents, available metadata, new system architecture, and future plans.

DATA FLOW IN SUPPORT OF IAG SERVICES



Today's global network of geodetic observing sites includes 421 GNSS receivers, 40 laser ranging sites, 44 VLBI stations, and 56 DORIS sites. The CDDIS provides data from these sites and higher-level products derived from these data.



An essential element to the operation of all of the IAG's international scientific services (the IGS, ILRS, IVS, and IDS) is their information system and archive components. These archives are the central source of data for the services' analysis communities as well as those products generated by the analysis centers for use by a broader user community. The CDDIS serves as the single data center resource from which the data from all these networks (IGS, ILRS, IVS, and IDS), and the resulting products derived from these data, may be accessed.

The IGS, as well as the ILRS, IVS, and IDS, makes use of a similar, distributed data flow structure for the transmission of information, data, and derived products from the observing stations to the user community. The structure for the services can be divided into the following components: Network Stations, Data Centers Analysis Centers, Analysis Center Coordinators, a Coordinating Center, and a Governing Body. Participants in these service activities collaborate at all levels to ensure consistency and timely delivery of data and products. GNSS (and laser ranging, VLBI, and DORIS) data are transmitted through the various levels shown in the figure to ultimately arrive at the analysis centers, combination centers, and general user community.

CDDIS REPORTING

IGS Tracking Network Status for 10-May-10 100510 10130 GPS Week 1583 Day 2 As of date: May 17 2010 10:27:21

Site	Dly (H)	No. Exp.	No. Obs.	Pts. Del.	%	Avg. MP1	Avg. MP2	Pos. Diff	No. Slps	V	Receiver Type	Antenna Type		Ant. Height	Marker Name			Marker Number	Type	RINEX Version	Dly (M)
abmf	10	25780	25640	136	99	0.41	0.51	0.04	23	1	TRIMBLE NETR5	TRM55971.00	NONE	0.0000	ABMF			97103M001	M	2.11	630
abpo	1	26268	25800	0	98	0.35	0.37	0.09	23	1	ASHTECH UZ-12	ASH701945G_M	SCIT	0.0083	ABPO			33302M001	G	2.11	14
ade1	20	25447	25133	0	98	0.43	0.41	0.04	4	1	ASHTECH Z-XII3	ASH700936B_M	SNOW	0.0000	ade1			50109S001	G	2.11	1253
ade2	20	25447	25071	1	98	0.42	0.41	0.04	5	1	ASHTECH Z-XII3	ASH700936B_M	SNOW	0.0000	ade2			50109S001	G	2.11	1253
adis	24	28500	22796	672	79	0.53	0.58	0.03	22	1	JPS LEGACY	TRM29659.00	NONE	0.0010	ADIS			31502M001	M	2	1490
aira	1	24806	22489	608	90	0.35	0.48	0.03	30	1	TRIMBLE 5700	TRM29659.00	DOME	0.0000	AIRA			21742S001	G	2.11	111
ajac	8	23674	23644	0	99	0.19	0.16	0.04	2	1	LEICA GRX1200GGPRO	LEIAT504GG	NONE	0.0000	AJAC			10077M005	M	2	511
albh	1	25563	25235	0	98	0.22	0.26	0.03	3	1	AOA BENCHMARK ACT	AOAD/M_T	SCIS	0.1000	albh	WCDA-ACP	927	40129M003	G	2.11	8
algo	1	24839	24730	0	99	0.22	0.23	0.07	1	1	AOA BENCHMARK ACT	AOAD/M_T	NONE	0.1000	ALGO	CACS-ACP	8831	40104M002	G	2.11	11
.																					
.																					
.																					
zamb																					
zeck	9	24586	24311	32	98	0.49	0.51	0.07	5	1	ASHTECH Z-XII3	ASH700936D_M	SNOW	0.0450	ZECK			12351M001	G	2.10	570
zim2	1	24423	24342	2	99	0.29	0.34	0.06	4	1	TRIMBLE NETR5	TRM59800.00	NONE	0.0000	ZIM2			14001M008	M	2.11	12
zimj	9	24432	24069	48	98	0.51	0.55	0.03	3	1	JPS LEGACY	JPSREGANT_SD_E	NONE	0.0770	ZIMJ			14001M006	M	2	541
zimm	1	24423	24177	197	98	0.38	0.34	0.07	47	1	TRIMBLE NETRS	TRM29659.00	NONE	0.0000	ZIMM			14001M004	G	2.1	8
zwe2	9	25076	24785	0	98	0.28	0.28	0.07	39	1	SEPT POLARX2	TPSCR3_GGD	NONE	0.1300	ZWE2			12330M003	M	2.11	541

Program: QC 2009Mar23 by UNAVCO run with elevation angle cutoff of 10 degrees

Field	Size	Type	Explanation
Site	4	char	Site name
Dly (H)	3	number	Delivery delay in hours
No. Exp.	5	number	Total number of observations expected
No. Obs.	5	number	Total number of observations in file
Pts. Del.	5	number	Total number of points deleted
%	3	number	Data collection percentage
Avg. MP1	4	number	Average L1 multipath (rounded to two decimal places)
Avg. MP2	4	number	Average L2 multipath (rounded to two decimal places)
Pos. Diff	4	number	RINEX vs QC point position difference (Km)
No. Slps	4	number	Number of detected slips
V	1	number	Version of the data file (set to 1 for initial delivery)
Receiver Type	20	char	Type of GPS receiver from RINEX header
Antenna Type	20	char	Type of GPS antenna from RINEX header
Ant. Height	6	number	Height of antenna from RINEX header
Marker Name	20	char	Marker name from RINEX header
Marker Number	10	char	Marker DOMES number from RINEX header
Type	1	char	GPS type
RINEX Version	7	char	RINEX Version number
Dly (M)	3	number	Delivery delay in minutes

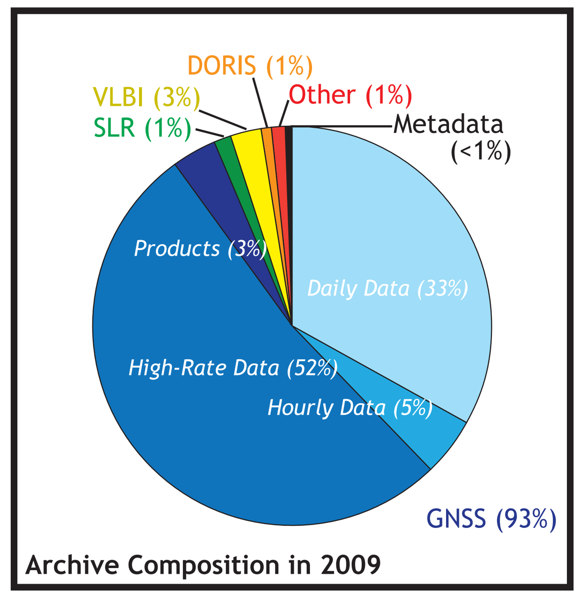
Hourly GNSS Archive Statistics -- 2010/05/01 to 2010/05/31

Mon.	00-04m	05-09m	19-29m	30-59m	01-24h	01-3d	3d-mis
abmf	----	----	90.80%	5.46%	3.74%	----	----
abpo	----	96.51%	3.36%	0.13%	----	----	----
adis	----	----	58.57%	2.39%	30.68%	8.37%	8.37%
aira	21.37%	1.61%	0.27%	75.94%	0.81%	----	----
albh	----	92.47%	2.28%	----	5.24%	----	----
algo	76.28%	12.13%	0.67%	0.81%	5.66%	4.45%	4.45%
alic	----	0.13%	6.72%	8.60%	73.52%	6.32%	6.32%
amc2	----	99.46%	0.54%	----	----	----	----
ankr	----	----	92.81%	3.52%	3.67%	----	----
areq	----	94.62%	2.69%	1.08%	1.61%	----	----
arev	----	----	----	----	----	----	----
artu	----	----	95.56%	0.13%	4.30%	----	----
aspa	26.94%	35.50%	0.16%	6.97%	22.50%	5.86%	5.86%
auck	----	----	0.13%	99.46%	0.40%	----	----
aukt	----	----	11.42%	27.28%	61.29%	----	----
.							
.							
.							
zeck	----	----	90.94%	2.96%	6.10%	----	----
zim2	99.73%	----	----	0.13%	0.13%	----	----
zimj	10.22%	0.27%	79.03%	0.40%	4.70%	5.38%	5.38%
zimm	99.73%	----	----	----	0.27%	----	----
zwe2	12.90%	0.54%	73.12%	4.30%	9.14%	----	----
Avg.	20.26%	23.59%	38.24%	7.31%	8.42%	1.55%	1.55%
Mon.	00-04m	05-09m	19-29m	30-59m	01-24h	01-3d	3d-mis

The CDDIS provides information on the contents of the data and product archive, tabulated by data type (e.g., daily data, orbit products, etc.). Files containing summary information, such as data holdings, temporal and spatial coverage, data quality, statistics, are also available to aid the user in data discovery.

The archive status files of CDDIS GNSS data holdings reflect timeliness of the data delivered as well as statistics on number of data points, cycle slips, and multipath. The user community can receive a snapshot of data availability and quality by viewing the contents of such a summary file as shown in the figure. The CDDIS staff has recently enhanced the format of these files as requested by the Infrastructure Committee. Software for generating the files is available to other data centers.

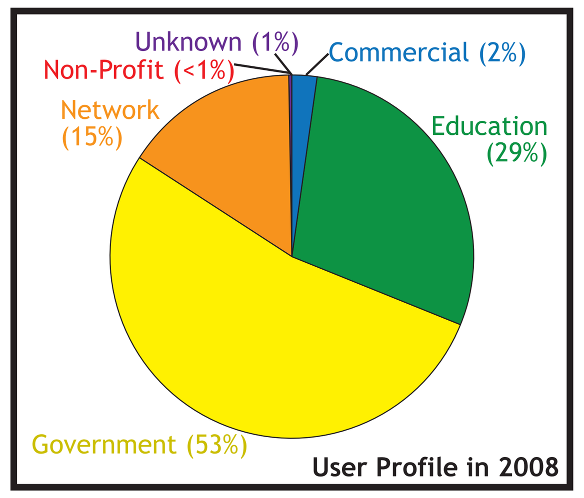
A new report summarizing data latency has recently been created at the CDDIS. This monthly/yearly report provides average latency statistics on the arrival of hourly data by site.



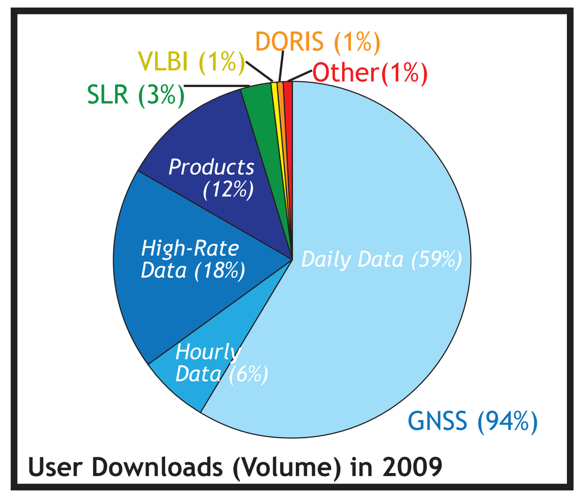
Archive Composition in 2009

The breakdown of the contents of the CDDIS archive by type in 2010. The majority of the online storage is devoted to the archive of GNSS data. The CDDIS GNSS archive consists of 19 years of daily, 30-second files (420 sites in 2010), 3 years hourly, 30-second data (280 sites in 2010),10 years of high-rate 1-second data (140 sites in 2010), and 18 years of GNSS product files.

Profile of users retrieving files (based on volume) from the CDDIS archive in 2008.



User Profile in 2008



User Downloads (Volume) in 2009

The breakdown of CDDIS archive downloads by volume for 2009 categorized by data/product type. A further breakdown by data format and processing level is shown for GNSS-related downloads.

CDDIS ARCHIVE PROCESSING

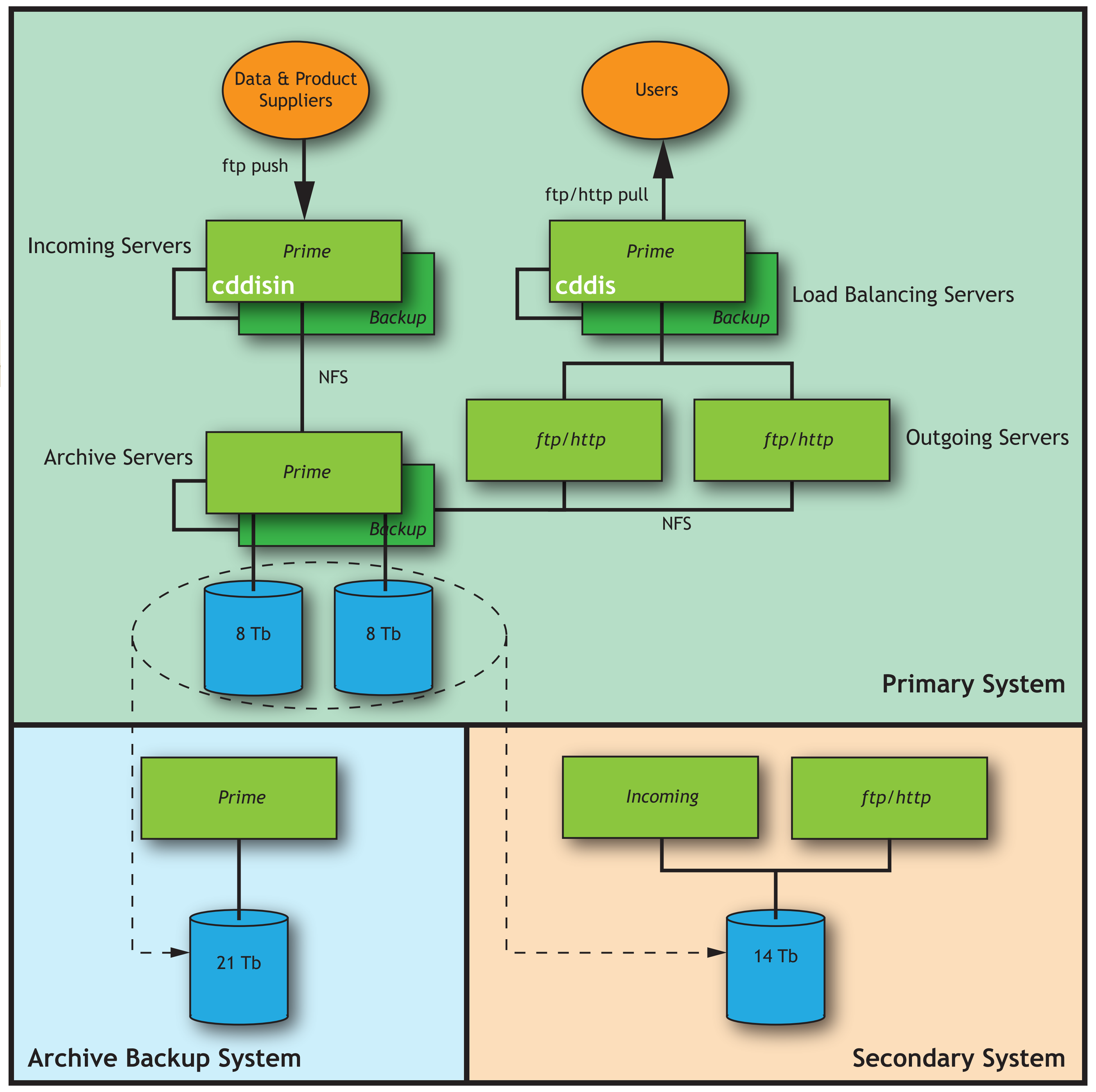
The update process for the CDDIS archive process can be divided into several structural components allowing for efficient and secure processing: deposit, operations, download, and archive support.

DEPOSIT: Suppliers of content for the archive (e.g., network tracking data from operational centers, products from analysis centers, etc.) transfer their data and product files to the CDDIS deposit or "incoming" disk location using ftp. These incoming accounts have limited privileges allowing users to only deposit files. In a few cases, the CDDIS will retrieve files for the archive from data/product sources. All suppliers access a server dedicated to receipt of incoming files.

OPERATIONS: All processing of incoming files takes place in the CDDIS operations area, which is accessible to internal users only. Software scans the deposit directories on pre-determined schedules dependent upon the type of incoming files and copies the files to temporary locations where their contents are validated for readability and integrity (format and content) and metadata are extracted and loaded into a relational database. Valid files are moved to the CDDIS archive.

DOWNLOAD: The CDDIS public archive is openly accessible to the scientific community through anonymous ftp and the web (future enhancements will permit http access to the CDDIS archive). It is the repository for all valid files provided by the operational/regional/global data centers, analysis centers, and analysis center coordinators. The structure of the archive follows conventions established within the services and thus is data type (i.e., GNSS, SLR, VLBI, or DORIS) dependent. All users access a separate computer system dedicated to serving files from the archive's disk farm.

ARCHIVE SUPPORT: A final portion of the CDDIS archive update process is devoted to utilizing extracted metadata to maintain supporting information, particularly files summarizing the contents of the download area, statistics on the timeliness of the incoming files, etc.



NEW CDDIS COMPUTER ENVIRONMENT

On June 21, 2010 the CDDIS transitioned operations to a new distributed server environment. This new configuration allows for efficient and timely processing of incoming files as well as enhanced system security by separating user/archive functions. Distinct servers handle incoming data and product files (server *cddisin.gsfc.nasa.gov*), outgoing ftp and http requests (server *cddis.gsfc.nasa.gov*), and archive operations to the RAID storage. Servers handle load balancing on incoming queries for files to host cddis.gsfc.nasa.gov. The archive server manages the RAID storage and its connections to the incoming and outgoing servers. Each server has a "hot spare" which can take over operations should a failure occur with the prime server. Additional RAID storage has been installed to bring the total available storage for the CDDIS archive to nearly 16 Tbytes, plus additional internal storage for processing and database applications. The CDDIS archive increases in size by approximately 1 Tbyte/year; the existing storage will accommodate the archive requirements for the near future. The CDDIS computer system also includes a secondary server for daily backup of the archive. Furthermore, two additional servers and RAID arrays will be set up in the next few months at another GSFC location to provide a complete backup server environment should access to the primary systems be disabled.

In addition to computer hardware changes, the CDDIS replaced its internal database management software (Oracle) with MySQL. This change required modification to database schemas, supporting software, and report queries.

FUTURE DEVELOPMENTS

METADATA UPDATES: Metadata can be thought of as "data about data". They describe what, where, when, and by whom a set of data or products were collected. Metadata are used to manage the data archive and aid in access to this archive.

Today, CDDIS software extracts metadata as part of its data validation and ingest process. The fields and format of these metadata are dependent upon data type and processing level. However, to support integration of information about IAG service data and product holdings into the GGOS portal, the CDDIS staff has begun development of a new metadata model to be used for all types of data and products archived in the system. The model will be based upon internationally-recognized metadata standards in cooperation with GGOS and the EOSDIS Clearing House (ECHO). Some of the parameters to be addressed in the new model include descriptions at various levels: collection (e.g., laser ranging normal points, IGS final orbits, etc.) granule (e.g., daily RINEX observation data for GODE on day 10152, IGS final clock solution for day 3 of week 1582, etc.) ancillary (e.g., station code GODE refers to a site at Greenbelt, MD, etc.).

Several standards have been developed for metadata. GGOS will base its metadata model on ISO 19511 and will provide core metadata information for all geodetic data sets from the contributing IAG services. Therefore, these services will need to provide metadata, through their data centers, following this model so that their data are discoverable in the GGOS portal. The model can be extended to include other metadata required by the individual services and or data centers.

USER INTERFACE ENHANCEMENTS: One area of improvement for the CDDIS is in discovery of the system's contents for both new users and for the existing user base. The average user of the CDDIS today accesses the contents of the archive through anonymous ftp by means of automated scripts. Analysts can use this method for data transfer because they are familiar with the structure of the CDDIS and thus know what files they require, their availability schedule, and where to find them within the online structure. However, new users of the CDDIS, both those familiar with space geodesy techniques as well as new research communities, would prefer a browsing interface to the archive contents. Therefore, the CDDIS will undertake the design of a web interface based search tool that queries the CDDIS metadata. Users will have the ability to specify search criteria based on temporal, spatial, target, site designation, and/or observation parameter in order to identify data and products of interest for download. Results of these queries will include a listing of sites (or other metadata) or data holdings satisfying the user input specifications. Such a user interface will also aid CDDIS staff in managing the contents of the archive.

OTHER UPDATES UNDER CONSIDERATION: Data centers, including the CDDIS, fight a continuous battle to manage an ever-increasing archive in an efficient manner. During our transition to the new server environment, the CDDIS staff realized that the large number of nested directories used in archiving some data sets, e.g., high-rate GNSS data, is not efficient for storage or access. However, there is a need to balance efficiency with user requirements. To that end, the CDDIS will study ways to simplify storage of selected data sets, such as the high-rate GNSS data. As an example, a day's worth of GNSS high-rate data could be consolidated by packaging all files into a single site's tar file, (e.g., gode1150_10d.tar.gz contains all high-rate observation data from GODE for day 10115). Users can then download a single file/day to retrieve all high-rate data for sites of interest. Similar consolidation could be performed on other data sets as is practical. Any changes would be implemented in the historic data directories (e.g., prior to 2009); operational changes would be coordinate with IGS Infrastructure Committee and Data Center Working Group.

The current compression scheme used in the IGS (and other IAG service infrastructures), UNIX compress, is inefficient and out of date. The IGS Data Center Working Group (DCWG) is currently reviewing options for implementing alternate compression software. Any change to compression software would involve not only the data centers but other components including the analysis centers, user community in general, and receiver manufacturers, many of whom implement compression in their firmware. However, it is reasonable to allow data centers to alter compression used in older data sets to realize a more efficient data storage. Currently, gzip and bzip2 are options under investigation to replace UNIX compress.

The CDDIS staff will look for feedback prior to implementing any significant changes such as these to the archive.

REFERENCES AND FURTHER READING

C. Noll, The crustal dynamics data information system: A resource to support scientific analysis using space geodesy, Advances in Space Research, Volume 45, Issue 12, DORIS: Scientific Applications in Geodesy and Geodynamics, 15 June 2010, Pages 1421-1440, ISSN 0273-1177, DOI: 10.1016/j.asr.2010.01.018.

C. Noll, Y. Bock, H. Habrich and A. Moore, "Development of data infrastructure to support scientific analysis for the International GNSS Service", Journal of Geodesy, Feb 2009, pages 309-325, DOI 10.1007/s00190-008-0245-6.

C. Noll, The Global GNSS, SLR, VLBI, and DORIS Networks and their Support of GGOS: IGS+ ILRS+IVS+IDS, Eos Trans. AGU, 89(53), Fall Meet. Suppl., Abstract G33A-066, 20008.

C. Noll, L. Soudarin, "On-line Resources Supporting the Data, Products, and Infrastructure for the International DORIS Service", Journal of Geodesy, Jan 2006, Pages 1 - 9, DOI 10.1007/s00190-006-0051-y.